

# From 'What' to 'Why': AI-Powered Anomaly Detection in Snowflake

## The Call Center Challenge



## The 3-Layer Solution in Snowflake

### Step 1: Data Foundation

Build a secure data foundation using Snowflake tables and views with synthetic call data.



## The Outcome: From Data to Actionable Insight

### A Unified BI Dashboard

The final dashboard combines metrics, charts, and AI-generated text for a complete picture.



# Snowflake Prototyping

Loy Sikdar  
2025

# End-to-End Anomaly Detection & AI-Powered Explanation in Snowflake

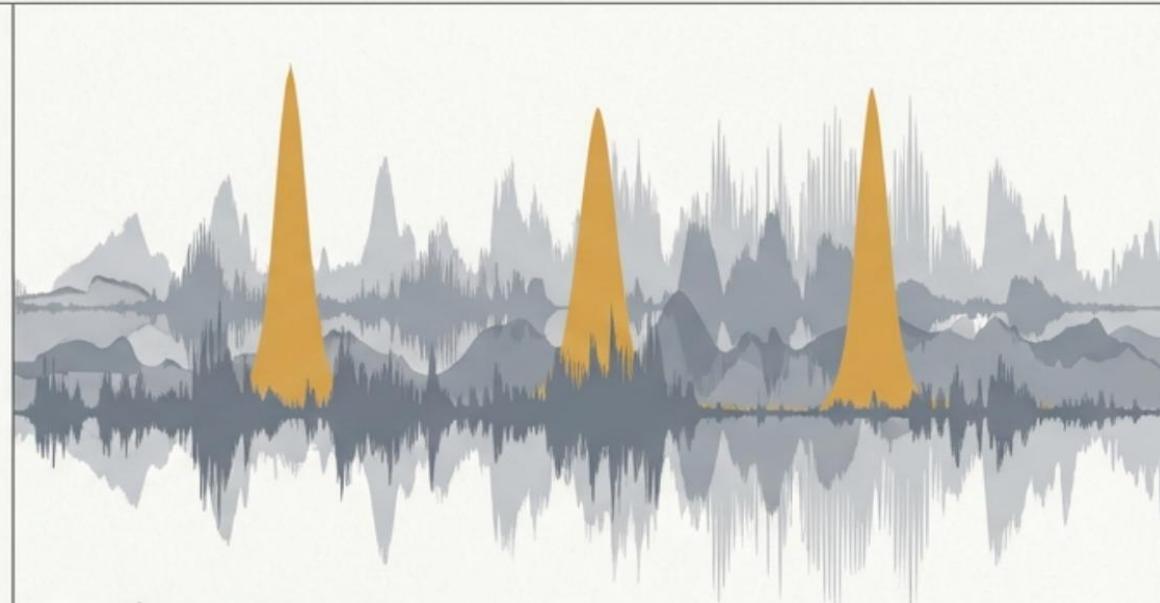
A technical walkthrough of a call center solution built with Snowpark, Cortex, and `mistral-large2`.



# The Supervisor's Challenge: Finding and Understanding Outliers at Scale

## Core Problem Description

Call centers generate massive volumes of interaction data. Within this data, critical outliers—such as calls with extremely long handle times, unusual knowledge management (KM) usage, or repeated escalations—are difficult to systematically identify and diagnose.



## Business Needs

- ✓ **Automated Detection:** A scalable method to flag anomalous calls without manual review.
- ✓ **Root Cause Understanding:** A clear way to understand *why* a call was flagged as anomalous.
- ✓ **Actionable Insights:** Guidance on what actions to take, from agent coaching to process improvement.

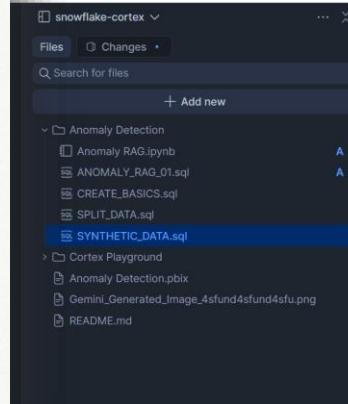
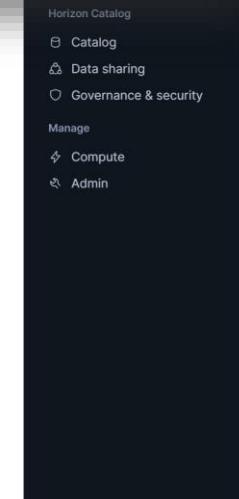
# Building the Foundation: A Synthetic Call Center Dataset

## Rationale

To develop and validate our models without using real customer data, we engineered a synthetic dataset simulating 12 months of activity (~100,000 call records).

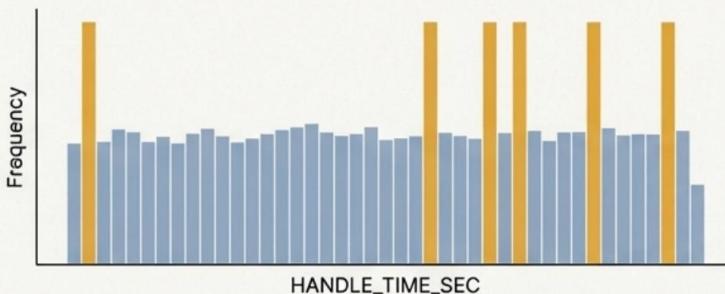
## Core Table

LAB\_CALL\_CENTER\_KM



# Engineering a Labeled Dataset for Model Development

## Anomaly Injection Strategy



We intentionally label ~10% of calls as anomalous with `ANOMALY\_REASON = 'ExtremeLongHandleTime'`.

These records have massively inflated `HANDLE\_TIME\_SEC` values compared to the 'Normal' population.

This creates a clean, labeled ground-truth set for evaluating model performance.

## Time-Based Data Splitting



The `DATASET\_SPLIT` column partitions the data chronologically using `CALL\_DATE`.

This structure supports realistic model training, validation, and simulated production scoring.

```
showflake-cortex > Anomaly Detection > SPLIT_DATA.sql
▶ ▾

1 USE WAREHOUSE ANOM_DETX_WH;
2 USE SCHEMA ANOM_DETX.AD_LABS;
3
4 ALTER TABLE LAB_CALL_CENTER_KM
5 ADD COLUMN IF NOT EXISTS DATASET_SPLIT STRING;
6
7
8 -- Label each row as TRAIN / TEST / PROD
9
10 UPDATE LAB_CALL_CENTER_KM
11 SET DATASET_SPLIT =
12 CASE
13     WHEN CALL_DATE < '2024-09-01' THEN 'TRAIN'
14     WHEN CALL_DATE < '2024-11-01' THEN 'TEST'
15     ELSE 'PROD'
16 END;
17
18 -- Create physical tables from the split
19 CREATE OR REPLACE TABLE LAB_CALL_CENTER_KM_TRAIN AS
20 SELECT *
21 FROM LAB_CALL_CENTER_KM
22 WHERE DATASET_SPLIT = 'TRAIN';
23
24 CREATE OR REPLACE TABLE LAB_CALL_CENTER_KM_TEST AS
25 SELECT *
26 FROM LAB_CALL_CENTER_KM
27 WHERE DATASET_SPLIT = 'TEST';
28
29 CREATE OR REPLACE TABLE LAB_CALL_CENTER_KM_PROD AS
30 SELECT *
31 FROM LAB_CALL_CENTER_KM
32 WHERE DATASET_SPLIT = 'PROD';
33
```

# Training the Detection Model with Snowpark and Scikit-learn

## Code to Data Flow



The entire ML workflow runs in Snowflake using Snowpark for Python. This allows us to bring the code to the data.

## Section 1: Feature Selection

We use a numeric feature set derived from LAB\_CALL\_CENTER\_KM, including:

- HANDLE\_TIME\_SEC
- KM\_TIME\_SEC
- AFTER\_CALL\_WORK\_SEC
- HOLD\_TIME\_SEC
- KM\_SEARCH\_COUNT
- AGENT\_TENURE\_MONTHS
- HOUR\_OF\_DAY

## Section 2: Algorithm

We use an **Isolation Forest**, an unsupervised algorithm well-suited for anomaly detection.

- **Key Parameter:** contamination=0.10, set to align with our known 10% anomaly injection rate.

The model is fit exclusively on the TRAIN data partition.

```
in <-- cell5
from sklearn.ensemble import IsolationForest
# contamination ~ expected anomaly rate (10% from your synthetic design)
iso = IsolationForest(
    n_estimators=200,
    contamination=0.10,
    random_state=42,
)
iso.fit(X_train)

atationForest(contamination=0.1, n_estimators=200, random_state=42)

down <-- cell13
Evaluate on TEST (using your IS_ANOMALY labels)

in <-- cell6
from sklearn.metrics import classification_report, roc_auc_score
# IsolationForest: -1 = anomaly, 1 = normal
test_preds = iso.predict(X_test)
test_scores = iso.decision_function(X_test) # higher = more normal

# Convert to 0/1 anomaly prediction
test_pred_is_anom = (test_preds == -1).astype(int)

print("Classification report (TEST):")
print(classification_report(y_test, test_pred_is_anom, digits=3))

# AUC using anomaly scores (flip sign so higher score = more anomalous)
test_auc = roc_auc_score(y_test, -test_scores)
print("ROC AUC (TEST):", round(test_auc, 3))

Classification report (TEST):
          precision    recall   f1-score   support
           0       0.901     0.905     0.903     14995
           1       0.129     0.124     0.126      1708
    accuracy                           0.825     16703
   macro avg       0.515     0.514     0.514     16703
weighted avg       0.822     0.825     0.823     16703
ROC AUC (TEST): 0.532
```

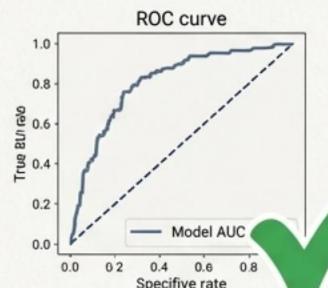
# Evaluating Performance and Scoring Production Data



## Evaluation on TEST Data

We score the 'TEST' partition and compare the model's predictions against the ground-truth 'IS\_ANOMALY' labels.

|                | Precision | Recall | F1-Score |
|----------------|-----------|--------|----------|
| 0              | 0.00      | 0.70   | 0.90     |
| 1              | 0.00      | 0.70   | 0.70     |
| Linear Score   | 0.85      | 0.88   |          |
| Standard Score | 0.50      | 0.50   |          |



Performance is measured using a classification report (precision, recall, F1) and ROC AUC, confirming the model effectively identifies the synthetic anomalies.



## Scoring PROD Data & Persisting Results



The validated model is applied to the 'PROD' data partition. The results, 'PRED\_IS\_ANOMALY' (0/1) and 'ANOMALY\_SCORE' (a continuous value), are written back to a new table:

'LAB\_CALL\_CENTER\_KM\_PROD\_SCORED'

```
predict(X_prod)
decision_function(X_prod) # higher = more normal

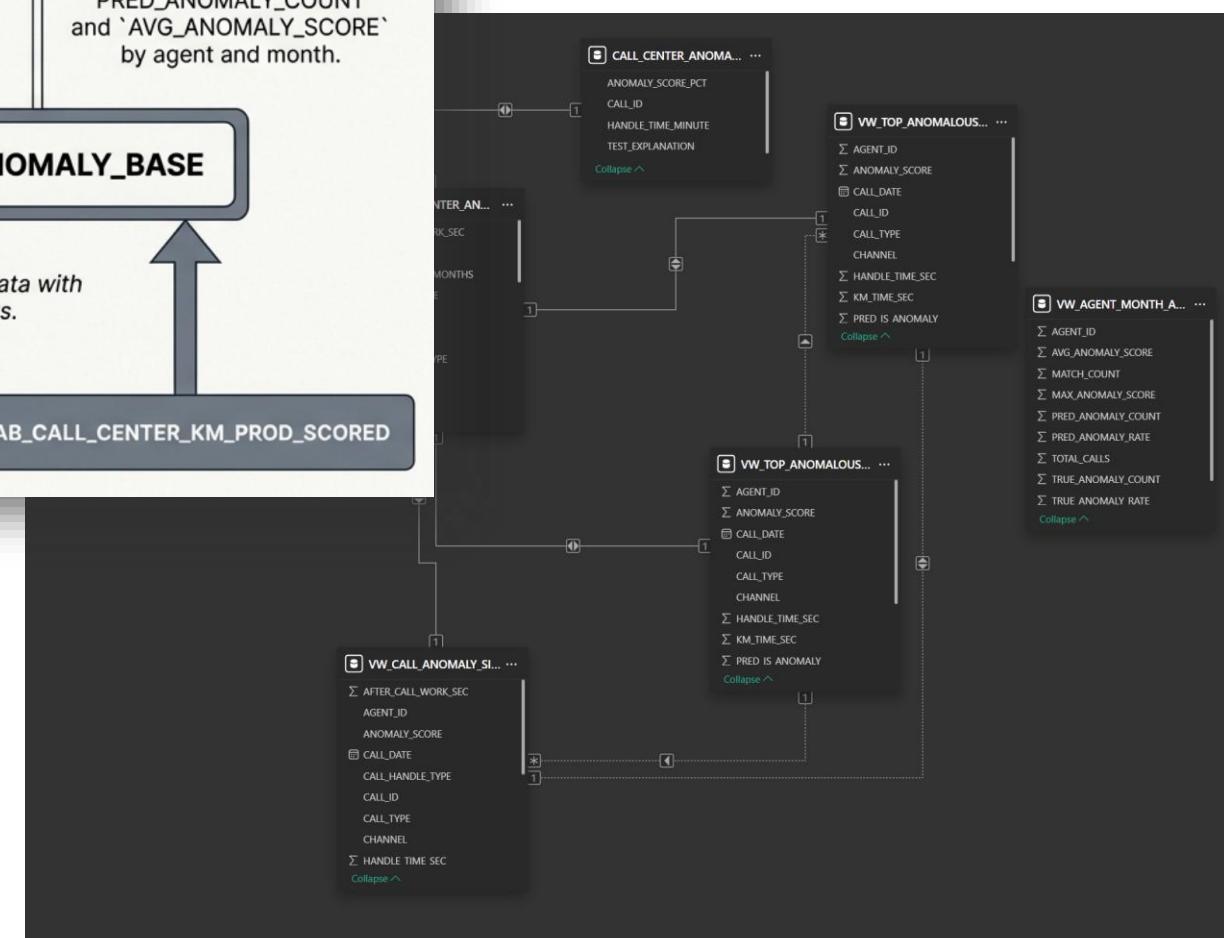
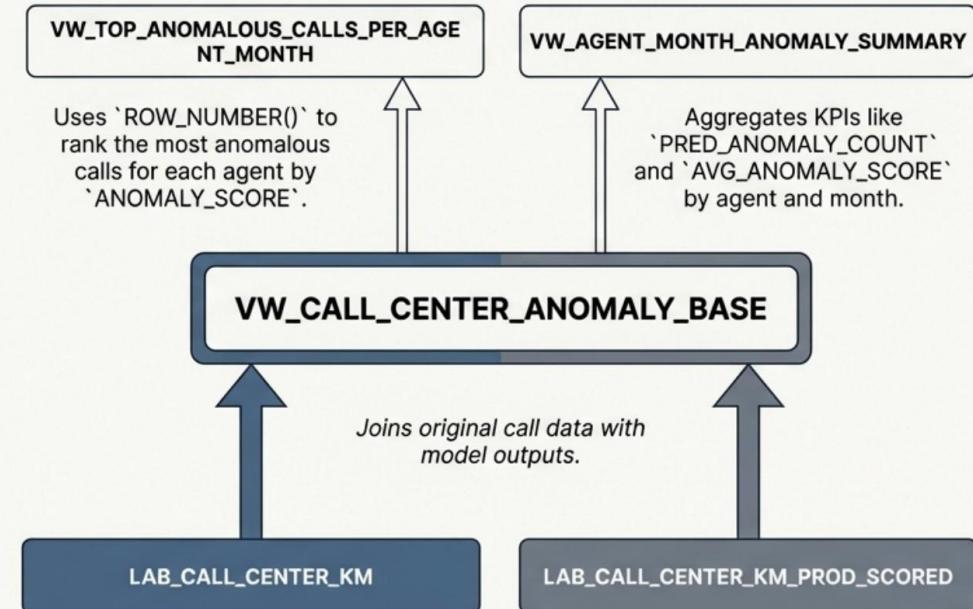
5 prod_pd["PRED_IS_ANOMALY"] = (prod_preds == -1).astype(int)
6 prod_pd["ANOMALY_SCORE"] = -prod_scores # higher = more anomalous
7
8 # Convert back to Snowpark DF
9 scored_prod_df = session.create_dataframe(prod_pd)
10
11 # Persist as a new table
12 scored_prod_df.write.save_as_table(
13     "LAB_CALL_CENTER_KM_PROD_SCORED",
14     mode="overwrite"
15 )
16
```

# Building a Semantic Layer with SQL Views for Simplified Analytics

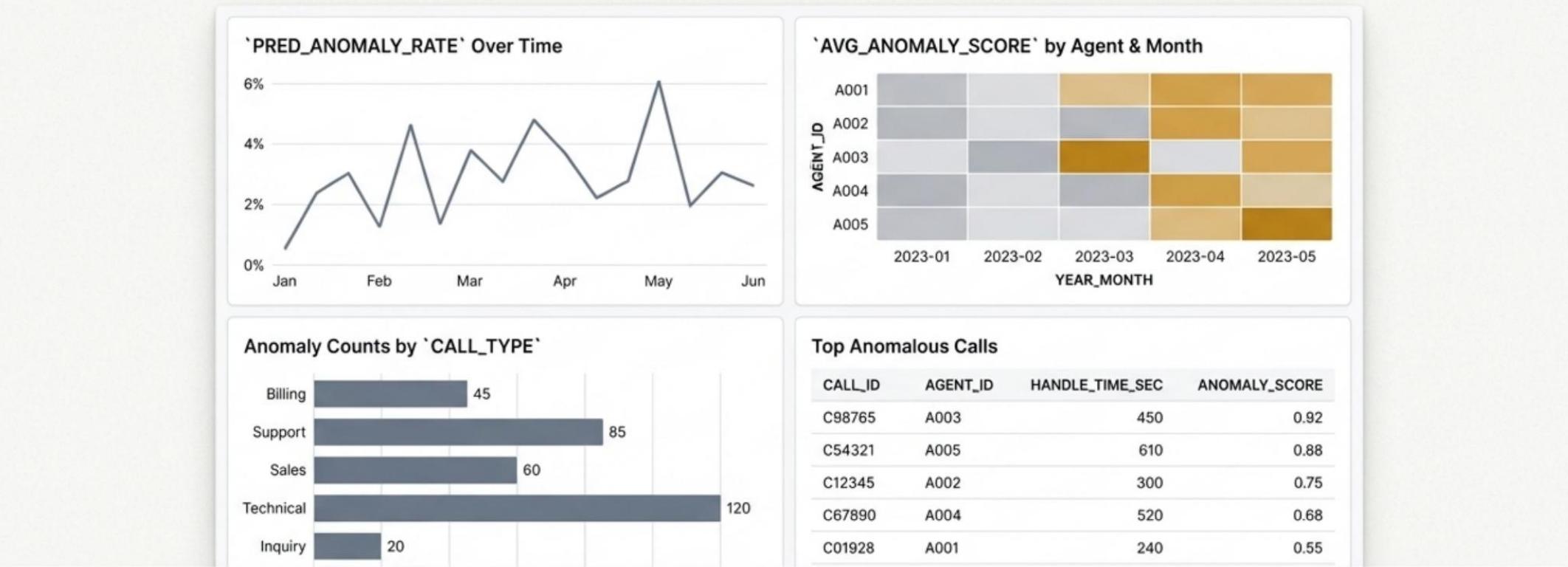
## Objective

To provide a clean, unified data source for BI tools and downstream analysis without exposing the complexity of underlying tables.

Source Sans Pro Regular



# Visualizing Anomalies: The Detection Dashboard



## What we know

We can now precisely identify which agents, call types, and time periods are experiencing the most anomalies.

## What we're missing

The quantitative data tells us *what* is anomalous, but not *why*. This requires manual investigation.

# From 'What' to 'Why': Introducing Snowflake Cortex for AI Explanations

## The Game-Changer

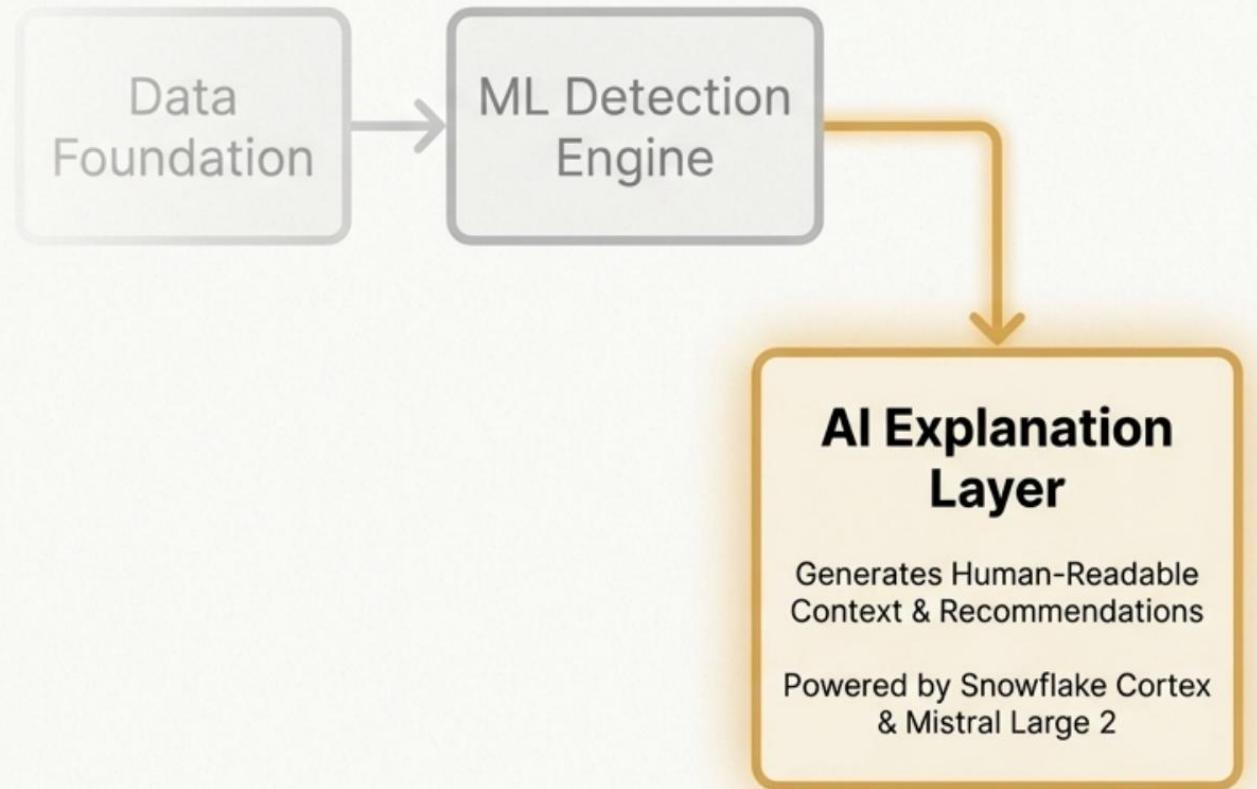
Snowflake Cortex provides built-in LLM capabilities directly via SQL functions, allowing us to generate narrative context without moving data out of the platform.

## Core Function

`SNOWFLAKE.CORTEX.COMPLETE` with the `mistral-large2` model.

## Objective

Convert raw metrics and anomaly scores into plain-language, human-readable explanations and recommendations.

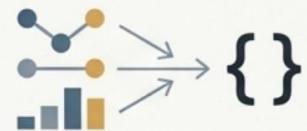


# Generating Per-Call Explanations with a Cortex Complete Function



## Step 1: Define a Persona

The prompt begins by instructing the model: "You are a senior call center performance analyst..."



## Step 2: Construct a Payload

For each anomalous call in `VW\_CALL\_CENTER\_ANOMALY\_BASE`, we use `OBJECT\_CONSTRUCT(...)` to create a JSON object containing key metrics (`HANDLE\_TIME\_SEC`, `KM\_TIME\_SEC`, `ANOMALY\_SCORE`, etc.).



## Step 3: Invoke the LLM

We pass this JSON payload to the `SNOWFLAKE.CORTEX.COMPLETE('mistral-large2', ...)` function.

## Expected Output

### Call ID: 12345-ABC

This call is anomalous due to an exceptionally high handle time (3200s), which is 5x the agent's average. This suggests a complex customer issue or procedural difficulty.

Coaching action: Review call recording with agent to identify knowledge gaps in handling this call type. Assess if KM articles need updating.

Database Explorer

Objects Data Products

Search

Filter

Tables 7

- CALL\_CENTER\_ANOMALY\_EXPLAINED
- CALL\_CENTER\_ANOMALY\_EXPLANATIONS
- LAB\_CALL\_CENTER.KM
- LAB\_CALL\_CENTER.KM.PROD
- LAB\_CALL\_CENTER.KM.PROD\_SCORED
- LAB\_CALL\_CENTER.KM.TEST
- LAB\_CALL\_CENTER.KM.TRAIN

Views 6

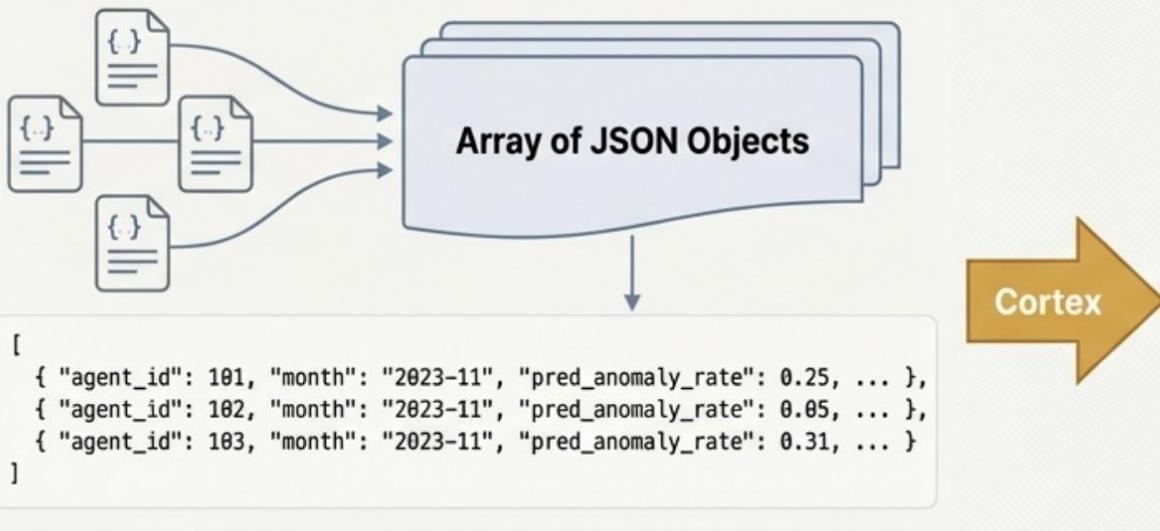
```
ANOMALY_EXPLANATION STRING  
);  
  
INSERT INTO CALL_CENTER_ANOMALY_EXPLANATIONS  
(  
    CALL_ID,  
    EXPLANATION_TS,  
    MODEL_NAME,  
    ANOMALY_EXPLANATION  
)  
SELECT  
    v.CALL_ID,  
    CURRENT_TIMESTAMP(),  
    'mistral-large2',  
    SNOWFLAKE.CORTEX.COMPLETE(  
        'mistral-large2',  
        'You are a senior call center performance analyst.'  
    );  
  
Given the metrics for one call that has been flagged as an anomaly, write:  
- a short explanation of why this call is unusual.  
- which signals stand out.  
- and one specific coaching or process recommendation for the supervisor.  
  
Limit to about 120 words.  
  
Call metrics (JSON):  
' {  
    TO_JSON(  
        OBJECT_CONSTRUCT(  
            'call_id', v.CALL_ID,  
            'agent_id', v.AGENT_ID,  
            'call_date', v.CALL_DATE,  
            'year_month', v.YEAR_MONTH,  
            'call_type', v.CALL_TYPE,  
            'call_handle_type', v.CALL_HANDLE_TYPE,  
            'channel', v.CHANNEL,  
            'handle_time_sec', v.HANDLE_TIME_SEC,  
            'km_time_sec', v.KM_TIME_SEC,  
            'after_call_work_sec', v.AFTER_CALL_WORK_SEC,  
            'hold_time_sec', v.HOLD_TIME_SEC,  
            'km_search_count', v.KM_SEARCH_COUNT,  
            'km_articles_viewed', v.KM_ARTICLES_VIEWED,  
            'anomaly_score', v.ANOMALY_SCORE,  
            'pred_is_anomaly', v.PRED_IS_ANOMALY,  
            'true_is_anomaly', v.TRUE_LABEL_IS_ANOMALY,  
            'true_anomaly_reason', v.TRUE_ANOMALY_REASON  
        )  
    ) AS ANOMALY_EXPLANATION  
FROM VW_CALL_CENTER_ANOMALY_BASE v  
WHERE v.PRED_IS_ANOMALY = 1
```

# Uncovering Systemic Patterns with Aggregate-Level Summaries

**Beyond Single Calls:** We can use Cortex to analyze trends across multiple data points by feeding it aggregated data.

## Example Use Case: Input

Individual agent summaries from  
'VW\_AGENT\_MONTH\_ANOMALY\_SUMMARY'



## Cortex Task

Analyze the provided data to identify agents most at risk, describe common patterns across their anomalies, and recommend systemic interventions.

## Generated Narrative Summary

### Key Observation:

Agents 101 and 103 consistently show the highest anomaly rates, primarily driven by long handle times on 'Billing Inquiry' calls.

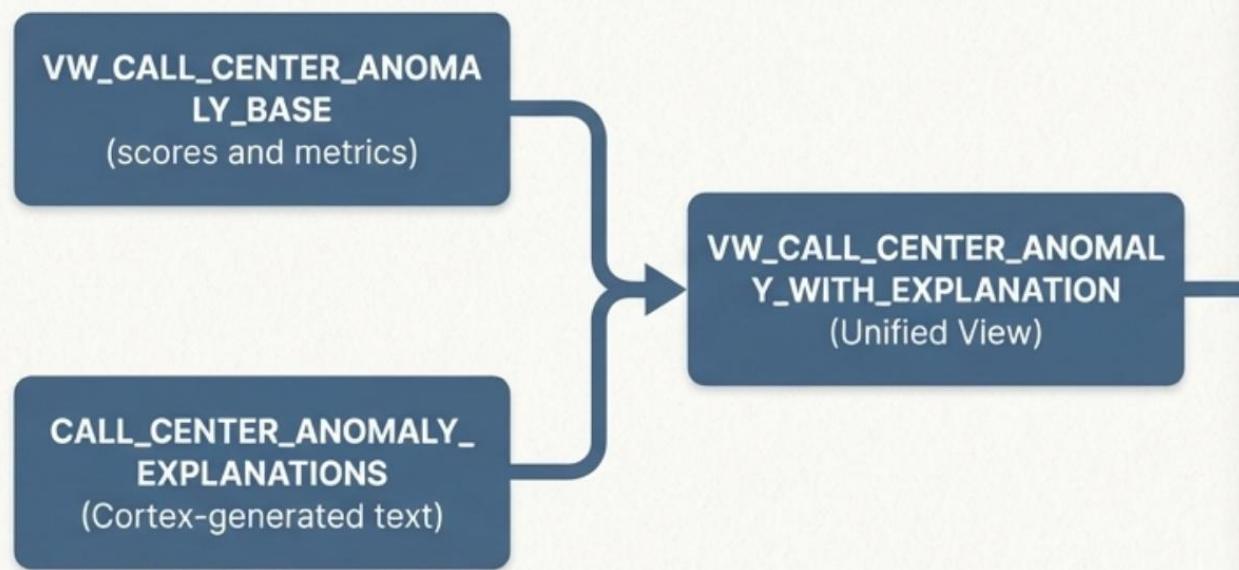
### Common Pattern:

These anomalies frequently occur during the last week of the month, suggesting a potential issue with end-of-month promotions or system performance.

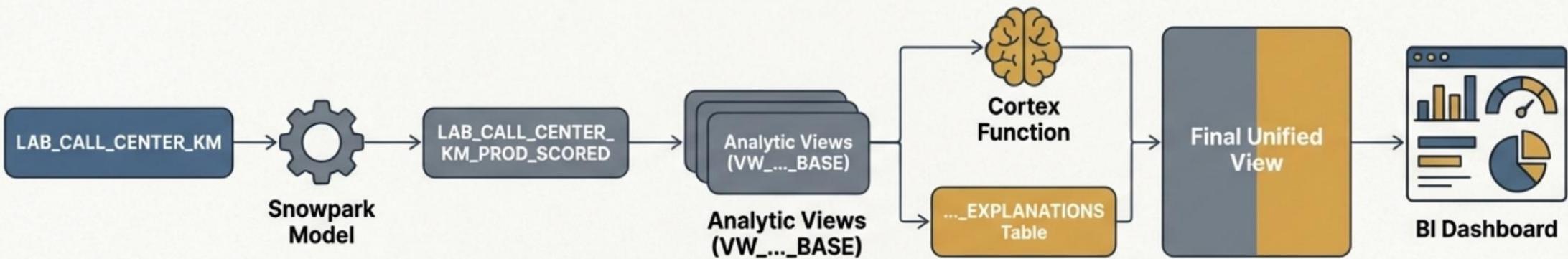
### Systemic Recommendation:

Propose a targeted training session for all agents on the 'Billing Inquiry' workflow and investigate the performance of the CRM tool during peak, end-of-month periods.

# The Complete Solution: Integrating AI Narratives into the BI Layer



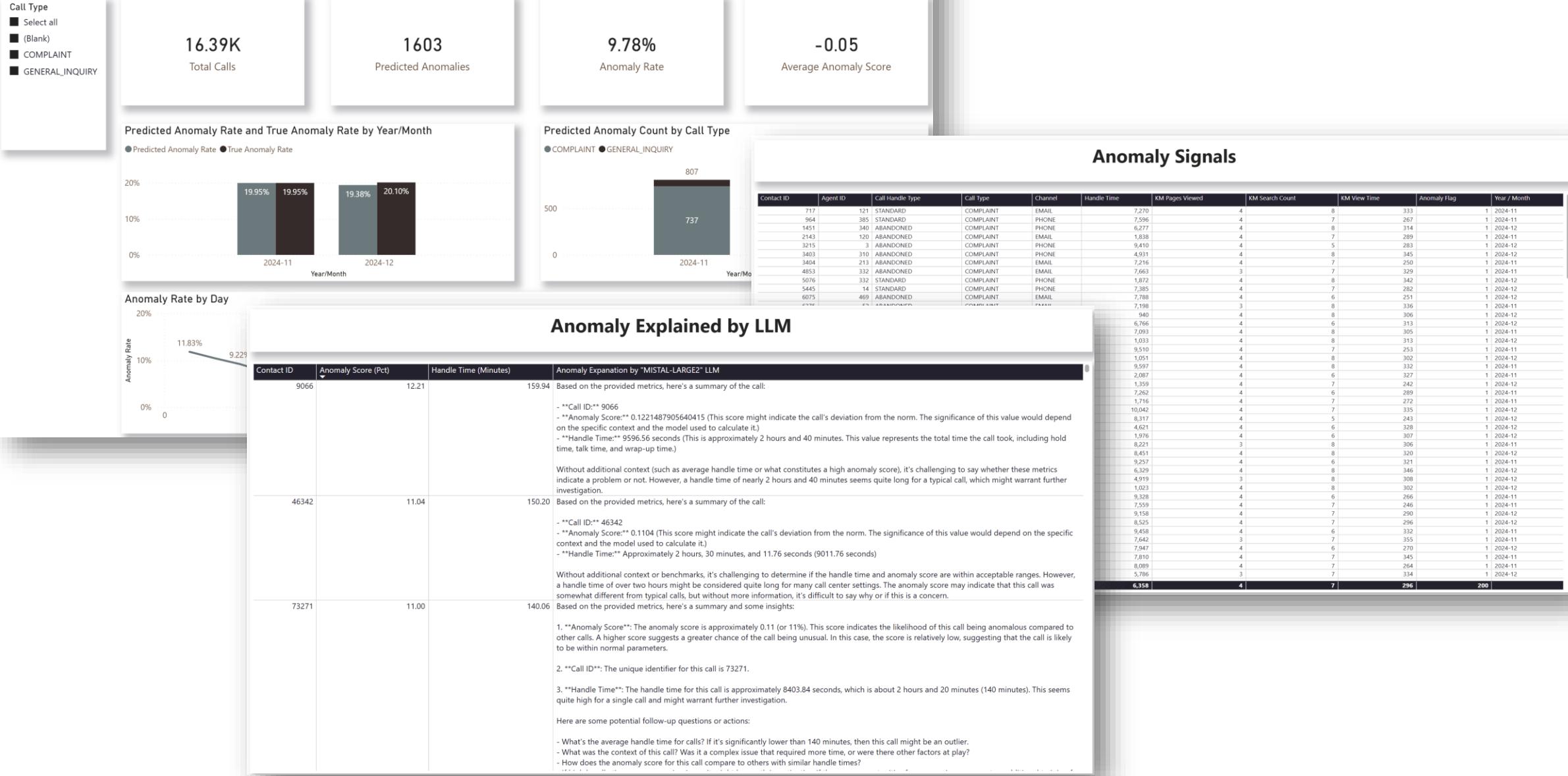
# The End-to-End Snowflake Workflow: From Data to Explained Insight



- 1. Data Generation:** Create a labeled synthetic dataset in Snowflake.
- 2. Model Training & Scoring:** Use Snowpark to build and apply an anomaly detection model.
- 3. Semantic Layer:** Construct SQL views to simplify data for analytics.
- 4. BI Visualization:** Identify what is anomalous in dashboards.
- 5. AI Explanation:** Use Cortex to generate narratives explaining why.

## Anomaly Detection - Handle Time and Time in Knowledge Management Pages

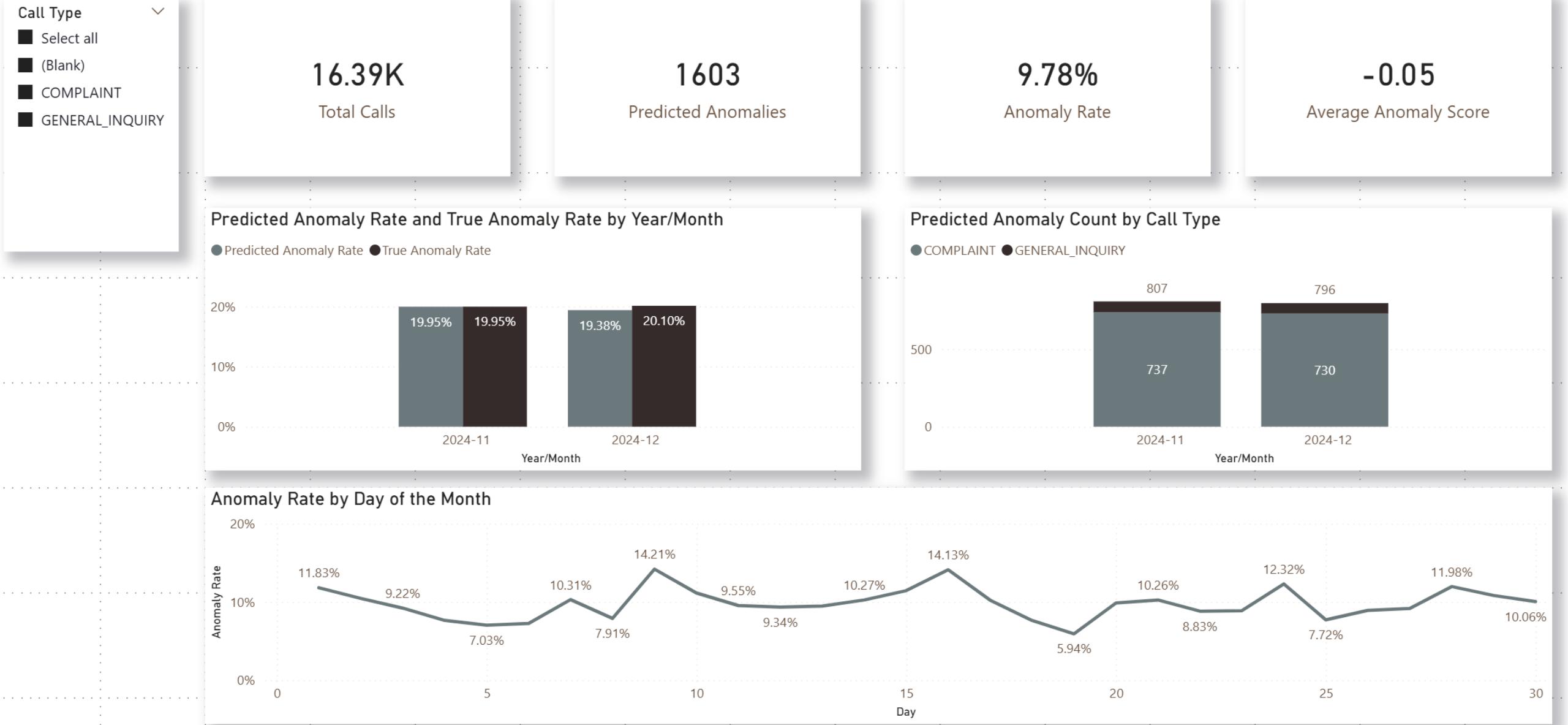
# Prototype Dashboard



**Anomaly Explained by LLM**

| Contact ID | Anomaly Score (Pct) | Handle Time (Minutes) | Anomaly Explanation by "MISTAL-LARGE2" LLM  |
|------------|---------------------|-----------------------|---|
| 9066       | 12.21               | 159.94                | <p>Based on the provided metrics, here's a summary of the call:</p> <ul style="list-style-type: none"> <li>**Call ID:** 9066</li> <li>**Anomaly Score:** 0.1221487905640415 (This score might indicate the call's deviation from the norm. The significance of this value would depend on the specific context and the model used to calculate it.)</li> <li>**Handle Time:** 9596.56 seconds (This is approximately 2 hours and 40 minutes. This value represents the total time the call took, including hold time, talk time, and wrap-up time.)</li> </ul> <p>Without additional context (such as average handle time or what constitutes a high anomaly score), it's challenging to say whether these metrics indicate a problem or not. However, a handle time of nearly 2 hours and 40 minutes seems quite long for a typical call, which might warrant further investigation.</p>   |
| 46342      | 11.04               | 150.20                | <p>Based on the provided metrics, here's a summary of the call:</p> <ul style="list-style-type: none"> <li>**Call ID:** 46342</li> <li>**Anomaly Score:** 0.1104 (This score might indicate the call's deviation from the norm. The significance of this value would depend on the specific context and the model used to calculate it.)</li> <li>**Handle Time:** Approximately 2 hours, 30 minutes, and 11.76 seconds (9011.76 seconds)</li> </ul> <p>Without additional context or benchmarks, it's challenging to determine if the handle time and anomaly score are within acceptable ranges. However, a handle time of over two hours might be considered quite long for many call center settings. The anomaly score may indicate that this call was somewhat different from typical calls, but without more information, it's difficult to say why or if this is a concern.</p>   |
| 73271      | 11.00               | 140.06                | <p>Based on the provided metrics, here's a summary and some insights:</p> <ol style="list-style-type: none"> <li>**Anomaly Score:** The anomaly score is approximately 0.11 (or 11%). This score indicates the likelihood of this call being anomalous compared to other calls. A higher score suggests a greater chance of the call being unusual. In this case, the score is relatively low, suggesting that the call is likely to be within normal parameters.</li> <li>**Call ID:** The unique identifier for this call is 73271.</li> <li>**Handle Time:** The handle time for this call is approximately 8403.84 seconds, which is about 2 hours and 20 minutes (140 minutes). This seems quite high for a single call and might warrant further investigation.</li> </ol> <p>Here are some potential follow-up questions or actions:</p> <ul style="list-style-type: none"> <li>What's the average handle time for calls? If it's significantly lower than 140 minutes, then this call might be an outlier.</li> <li>What was the context of this call? Was it a complex issue that required more time, or were there other factors at play?</li> <li>How does the anomaly score for this call compare to others with similar handle times?</li> </ul> |

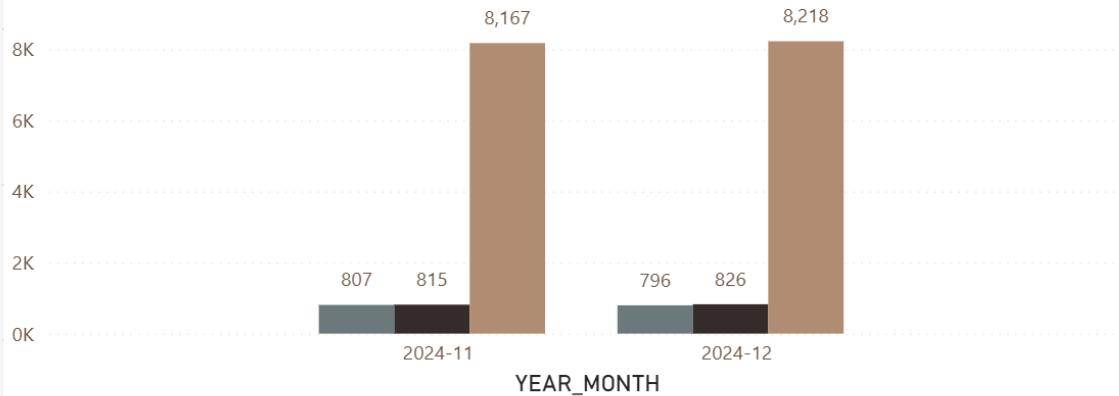
# Anomaly Detection - Handle Time and Time in Knowledge Management Pages



# Anomaly Detection - Patterns

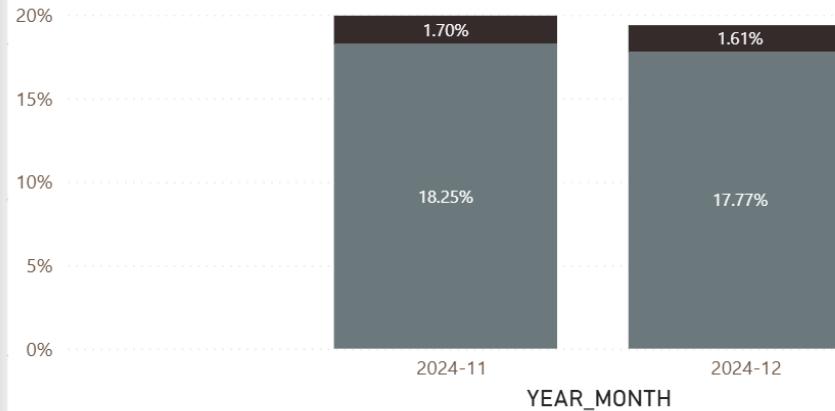
Predicted vs True Anomaly Count & Total Calls

● Predicted Anomaly Count ● True Anomaly Count ● Total Calls



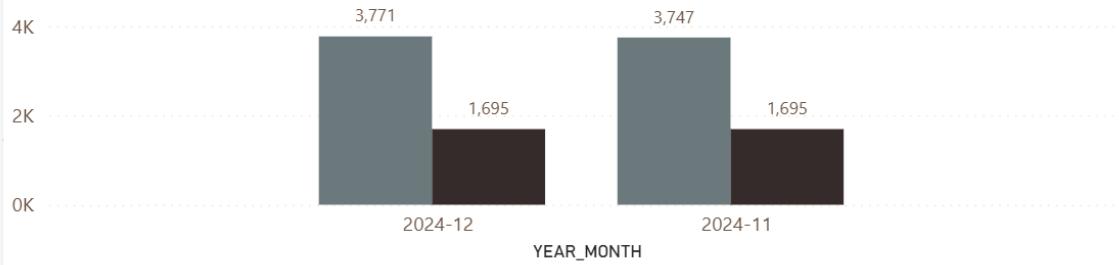
Anomaly Rate by Call Type by Month

● COMPLAINT ● GENERAL\_INQUIRY



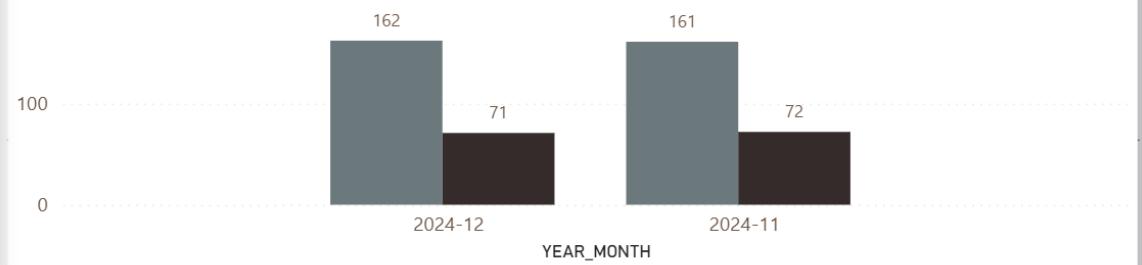
Avg Handle Time by Call Type by Month

● COMPLAINT ● GENERAL\_INQUIRY



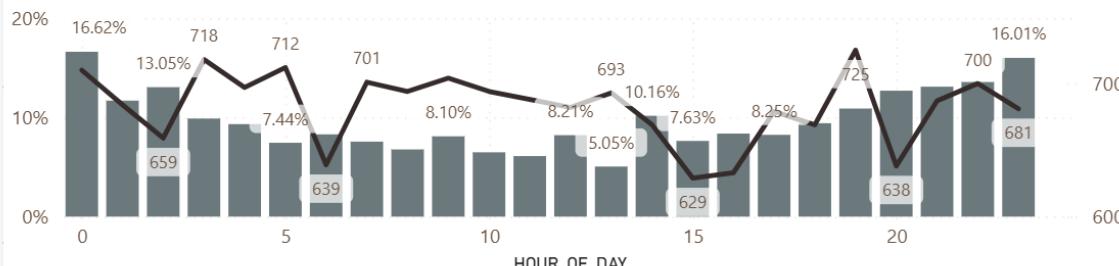
Avg Time in KM by Call Type by Month

● COMPLAINT ● GENERAL\_INQUIRY



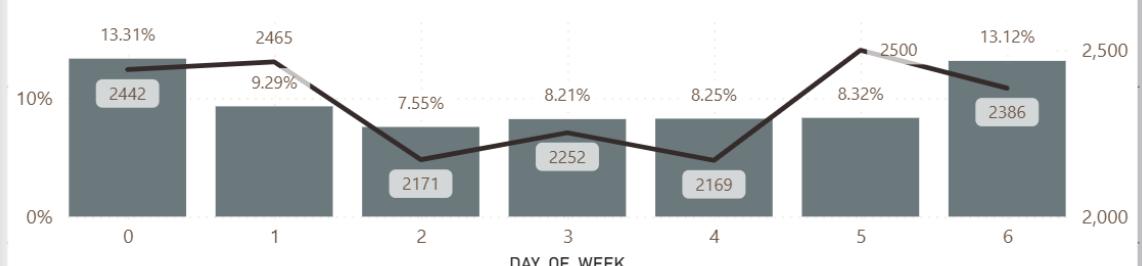
Call Count vs Anomaly Rate by Hour of the Day

● Anomaly Rate ● Count of CALL\_ID



Call Count vs Anomaly Rate by Day of the Week

● Anomaly Rate ● Count of CALL\_ID



| Complaint Analysis Dashboard - Q4 2023 |          |                  |           |              |             |                 |                 |                 |              |              |  |
|--|----------|------------------|-----------|--------------|-------------|-----------------|-----------------|-----------------|--------------|--------------|--|
| Contact Data                           |          |                  |           | Call Metrics |             |                 |                 | Support Tickets |              |              |  |
| Contact ID                             | Agent ID | Call Handle Type | Call Type | Channel      | Handle Time | KM Pages Viewed | KM Search Count | KM View Time    | Anomaly Flag | Year / Month |  |
| 717                                    | 121      | STANDARD         | COMPLAINT | EMAIL        | 7,270       | 4               | 8               | 333             | 1            | 2024-11      |  |
| 964                                    | 385      | STANDARD         | COMPLAINT | PHONE        | 7,596       | 4               | 7               | 267             | 1            | 2024-11      |  |
| 1451                                   | 340      | ABANDONED        | COMPLAINT | PHONE        | 6,277       | 4               | 8               | 314             | 1            | 2024-12      |  |
| 2143                                   | 120      | ABANDONED        | COMPLAINT | EMAIL        | 1,838       | 4               | 7               | 289             | 1            | 2024-11      |  |
| 3215                                   | 3        | ABANDONED        | COMPLAINT | PHONE        | 9,410       | 4               | 5               | 283             | 1            | 2024-12      |  |
| 3403                                   | 310      | ABANDONED        | COMPLAINT | PHONE        | 4,931       | 4               | 8               | 345             | 1            | 2024-12      |  |
| 3404                                   | 213      | ABANDONED        | COMPLAINT | EMAIL        | 7,216       | 4               | 7               | 250             | 1            | 2024-11      |  |
| 4853                                   | 332      | ABANDONED        | COMPLAINT | EMAIL        | 7,663       | 3               | 7               | 329             | 1            | 2024-11      |  |
| 5076                                   | 332      | STANDARD         | COMPLAINT | PHONE        | 1,872       | 4               | 8               | 342             | 1            | 2024-12      |  |
| 5445                                   | 14       | STANDARD         | COMPLAINT | PHONE        | 7,385       | 4               | 7               | 282             | 1            | 2024-12      |  |
| 6075                                   | 469      | ABANDONED        | COMPLAINT | EMAIL        | 7,788       | 4               | 6               | 251             | 1            | 2024-12      |  |
| 6375                                   | 52       | ABANDONED        | COMPLAINT | EMAIL        | 7,198       | 3               | 8               | 336             | 1            | 2024-11      |  |
| 6714                                   | 368      | STANDARD         | COMPLAINT | PHONE        | 940         | 4               | 8               | 306             | 1            | 2024-12      |  |
| 6884                                   | 182      | STANDARD         | COMPLAINT | EMAIL        | 6,766       | 4               | 6               | 313             | 1            | 2024-12      |  |
| 7494                                   | 412      | STANDARD         | COMPLAINT | EMAIL        | 7,093       | 4               | 8               | 305             | 1            | 2024-11      |  |
| 7851                                   | 213      | ABANDONED        | COMPLAINT | PHONE        | 1,033       | 4               | 8               | 313             | 1            | 2024-12      |  |
| 8500                                   | 220      | ABANDONED        | COMPLAINT | PHONE        | 9,510       | 4               | 7               | 253             | 1            | 2024-11      |  |
| 8927                                   | 165      | ABANDONED        | COMPLAINT | PHONE        | 1,051       | 4               | 8               | 302             | 1            | 2024-12      |  |
| 9066                                   | 348      | STANDARD         | COMPLAINT | EMAIL        | 9,597       | 4               | 8               | 332             | 1            | 2024-11      |  |
| 9432                                   | 185      | STANDARD         | COMPLAINT | PHONE        | 2,087       | 4               | 6               | 327             | 1            | 2024-11      |  |
| 9578                                   | 213      | STANDARD         | COMPLAINT | PHONE        | 1,359       | 4               | 7               | 242             | 1            | 2024-12      |  |
| 9920                                   | 441      | STANDARD         | COMPLAINT | EMAIL        | 7,262       | 4               | 6               | 289             | 1            | 2024-11      |  |
| 10353                                  | 73       | STANDARD         | COMPLAINT | PHONE        | 1,716       | 4               | 7               | 272             | 1            | 2024-11      |  |
| 10652                                  | 461      | STANDARD         | COMPLAINT | EMAIL        | 10,042      | 4               | 7               | 335             | 1            | 2024-12      |  |
| 10931                                  | 13       | ABANDONED        | COMPLAINT | PHONE        | 8,317       | 4               | 5               | 243             | 1            | 2024-12      |  |
| 10953                                  | 97       | ABANDONED        | COMPLAINT | EMAIL        | 4,621       | 4               | 6               | 328             | 1            | 2024-12      |  |
| 11803                                  | 298      | ABANDONED        | COMPLAINT | PHONE        | 1,976       | 4               | 6               | 307             | 1            | 2024-12      |  |
| 11820                                  | 453      | STANDARD         | COMPLAINT | PHONE        | 8,221       | 3               | 8               | 306             | 1            | 2024-11      |  |
| 12254                                  | 360      | STANDARD         | COMPLAINT | PHONE        | 8,451       | 4               | 8               | 320             | 1            | 2024-12      |  |
| 13689                                  | 273      | STANDARD         | COMPLAINT | PHONE        | 9,257       | 4               | 6               | 321             | 1            | 2024-11      |  |
| 15497                                  | 205      | ABANDONED        | COMPLAINT | PHONE        | 6,329       | 4               | 8               | 346             | 1            | 2024-12      |  |
| 15527                                  | 195      | STANDARD         | COMPLAINT | EMAIL        | 4,919       | 3               | 8               | 308             | 1            | 2024-12      |  |
| 16261                                  | 303      | ABANDONED        | COMPLAINT | PHONE        | 1,023       | 4               | 8               | 302             | 1            | 2024-12      |  |
| 16392                                  | 251      | STANDARD         | COMPLAINT | EMAIL        | 9,328       | 4               | 6               | 266             | 1            | 2024-11      |  |
| 16606                                  | 317      | STANDARD         | COMPLAINT | EMAIL        | 7,559       | 4               | 7               | 246             | 1            | 2024-11      |  |
| 17823                                  | 392      | ABANDONED        | COMPLAINT | EMAIL        | 9,158       | 4               | 7               | 290             | 1            | 2024-12      |  |
| 18217                                  | 332      | STANDARD         | COMPLAINT | EMAIL        | 8,525       | 4               | 7               | 296             | 1            | 2024-12      |  |
| 18402                                  | 423      | STANDARD         | COMPLAINT | PHONE        | 9,458       | 4               | 6               | 332             | 1            | 2024-11      |  |
| 19395                                  | 179      | STANDARD         | COMPLAINT | PHONE        | 7,642       | 3               | 7               | 355             | 1            | 2024-11      |  |
| 19910                                  | 406      | ABANDONED        | COMPLAINT | PHONE        | 7,947       | 4               | 6               | 270             | 1            | 2024-12      |  |
| 20873                                  | 259      | STANDARD         | COMPLAINT | EMAIL        | 7,810       | 4               | 7               | 345             | 1            | 2024-11      |  |
| 22437                                  | 279      | STANDARD         | COMPLAINT | PHONE        | 8,089       | 4               | 7               | 264             | 1            | 2024-11      |  |
| 24224                                  | 352      | STANDARD         | COMPLAINT | PHONE        | 5,786       | 3               | 7               | 334             | 1            | 2024-12      |  |
| Total                                  |          |                  |           |              | 6,358       | 4               | 7               | 296             | 200          |              |  |

# Anomaly Explained by LLM

| Contact ID | Anomaly Score (Pct) | Handle Time (Minutes) | Anomaly Explanation by "MISTAL-LARGE2" LLM   |
|------------|---------------------|-----------------------|--|
|            |                     |                       | <p>time, talk time, and wrap-up time.)</p> <p>Without additional context or benchmarks, it's challenging to determine if this call was unusually long or had other anomalies. However, a handle time of over 2.5 hours is generally quite high for a typical call center interaction, suggesting that this call might have involved complex issues or unusual circumstances.</p>   |
| 66995      | 10.45               | 149.58                | <p>Based on the provided metrics, here's a summary of the call:</p> <ul style="list-style-type: none"><li>- <b>Call ID</b>: 66995</li><li>- <b>Anomaly Score</b>: 0.1045 (This score might indicate the call's deviation from the norm. The significance of this value would depend on the specific context and the model used to calculate it.)</li><li>- <b>Handle Time</b>: 8974.76 seconds (This is approximately 2 hours and 29 minutes. This seems quite high for a typical call, suggesting that the call was unusually long.)</li></ul> <p>Given the long handle time, it might be worth investigating this call further to understand why it took so long and whether any issues need to be addressed. The anomaly score also suggests that this call was somewhat unusual compared to others.</p>  |
| 10652      | 10.22               | 167.37                | <p>Based on the provided metrics, here's a summary of the call:</p> <ul style="list-style-type: none"><li>- <b>Call ID</b>: 10652</li><li>- <b>Anomaly Score</b>: 0.102175345122043 (This score might indicate the call's deviation from the norm. The scale and significance of this score would depend on the specific model and context used.)</li><li>- <b>Handle Time</b>: 10042.4 seconds (approximately 2 hours and 47 minutes)</li></ul> <p>The handle time seems quite high, which might be indicative of a complex issue or an unusual situation during the call. The anomaly score also suggests that this call was somewhat different from typical calls. To better understand this, you might want to review the call details or compare it with other calls.</p>   |
| 97061      | 9.98                | 140.53                | <p>Based on the provided metrics, here's a summary of the call:</p> <ul style="list-style-type: none"><li>- <b>Call ID</b>: 97061</li><li>- <b>Anomaly Score</b>: 0.09977 (This score might indicate the call's deviation from the norm. A higher score might suggest an unusual call duration or content.)</li><li>- <b>Handle Time</b>: 8431.56 seconds (This is approximately 2 hours and 20 minutes. This seems quite long for a single call and could be a reason for the high anomaly score.)</li></ul> <p>Given the long duration, it might be worth reviewing this call to understand why it took so long and if there were any issues or complexities that led to this. It could be a unique case that required extensive discussion, or it might indicate a problem that needs addressing, such as agent training, process improvement, or technical issues.</p> |
| 40667      | 9.83                | 18.11                 | <p>Based on the provided metrics, here's a summary of the call:</p>  |

# From 'What' to 'Why': AI-Powered Anomaly Detection in Snowflake

## The Call Center Challenge



### Finding Outliers at Scale is Difficult

Call centers generate massive data, making it hard to manually find critical outliers.

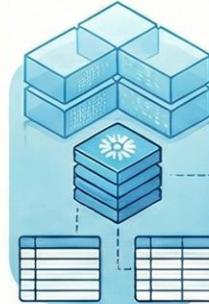
### Supervisors Need More Than Just Alerts

They need automated detection, root cause understanding, and clear, actionable insights.

## The 3-Layer Solution in Snowflake

### Step 1: Data Foundation

Build a secure data foundation using Snowflake tables and views with synthetic call data.



### Step 2: ML Detection Engine (The "What")

Build a secure data foundation using Snowflake tables and views with synthetic call data.



### Step 3: AI Explanation Layer (The "Why")

Use Snowflake Cortex to generate human-readable explanations for each flagged anomaly.



## The Outcome: From Data to Actionable Insight

### A Unified BI Dashboard

The final dashboard combines metrics, charts, and AI-generated text for a complete picture.



### Systemic Pattern Recognition

AI can also summarize trends across many calls to recommend stands, and manode improvements.

- This calls summarize trends improvement won the most callonized winnurs to recommend systemic systemic improvements.



# Value Realized and Future Roadmap

This solution demonstrates a complete, secure, and efficient workflow entirely within Snowflake, transforming raw data into not just detected anomalies, but understood and actionable insights.



## Automation

Schedule daily scoring and explanation jobs using Snowflake tasks.



## Model Expansion

Incorporate additional anomaly types, such as unusual KM usage patterns or abnormal call escalations.



## Operational Integration

Connect the insights directly to workforce management or QA platforms to trigger and track follow-up actions.

# Features and Capabilities needed for the prototype

## A. Compute & Language Capabilities (Snowpark)

- Feature:** Snowpark for Python enabled on specific warehouses.
- Privilege:** Ability to run Python user-defined functions (UDFs) and stored procedures.
- Justification:** We need to run data science workflows (using libraries like *scikit-learn* and *pandas*) directly where the data lives. This is required to train models (e.g., *Isolation Forests*) without the security risk and latency of exporting data to external servers

## B. Generative AI & LLM Access (Cortex)

- Feature:** Snowflake Cortex functions (specifically SNOWFLAKE.CORTEX.COMPLETE).
- Privilege:** Access to use Large Language Models (specifically mistral-large2) within SQL queries.
- Justification:** This allows us to convert raw numerical scores into human-readable narratives. We can programmatically generate explanations for business anomalies (e.g., "Why was this call flagged?") directly within the database

## C. Data Engineering Independence (Sandbox Schema)

- Feature/Priviledge:** CREATE STORED PROCEDURE, and CREATE TASK privileges within a designated "Lab" or "Sandbox" schema (e.g., ANOM\_DETX.AD\_LABS).
- Justification:** Logic Encapsulation and Automated scheduled